

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problems Mailbox.**

**THIS PAGE BLANK (USPTO)**

88/2858



33

(19) BUNDESREPUBLIK  
DEUTSCHLAND



DEUTSCHES  
PATENTAMT

(12) Patentschrift  
(10) DE 195 26 954 C 1

(51) Int. Cl. 6:  
**H 03 H 21/00**  
G 06 F 17/18

DE 195 26 954 C 1

(21) Aktenzeichen: 195 26 954.3-31  
(22) Anmeldetag: 24. 7. 95  
(43) Offenlegungstag: —  
(45) Veröffentlichungstag  
der Patenterteilung: 18. 4. 96

Innerhalb von 3 Monaten nach Veröffentlichung der Erteilung kann Einspruch erhoben werden

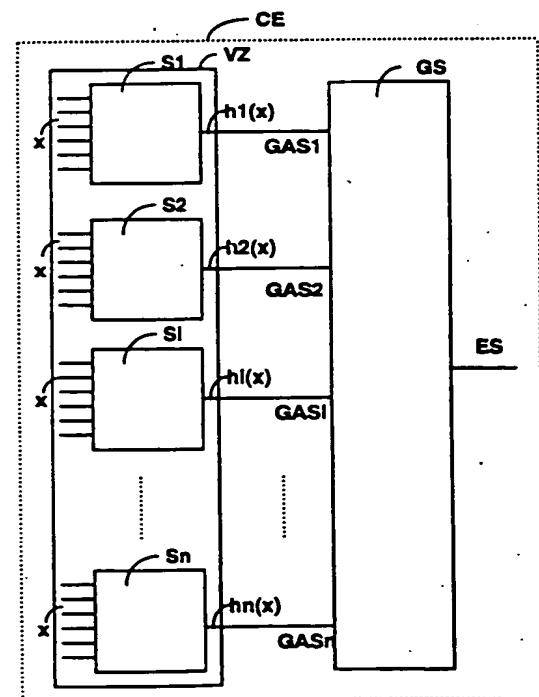
(73) Patentinhaber:  
Siemens AG, 80333 München, DE

(72) Erfinder:  
Tresp, Volker, Dr., 80539 München, DE; Taniguchi,  
Michiaki, 81371 München, DE

(56) Für die Beurteilung der Patentfähigkeit  
in Betracht gezogene Druckschriften:  
HAYKIN, Simon: Adaptive Filter Theory, 2nd ed.,  
Englewood Cliffs, New Jersey, Prentice-Hall,  
Inc. 1991, S.31-38 ISBN 0-13-013236-5;  
ROJAS, Raul: Theorie der neuronalen Netze, Berlin,  
Springer-Verlag, 1993, S.29-34 und 73-75  
ISBN 3-540-56353-9;

(54) Verfahren zur rechnergestützten Kombination einer Vielzahl von Schätzern, die auf statistischen Verfahren beruhen, insbesondere von neuronalen Netzen, zu einem Gesamtschätzer

(55) Ausgangssignale (ASi) einzelner Schätzer (Si) einer Vielzahl (VZ) von Schätzern (Si) werden gewichtet (GASi) und einem Gesamtschätzer (GS) zugeführt, der ein Ergebnissignal (ES(xg)) bildet.  
Zur Gewichtung werden Gewichtsfunktionen ( $h_i(x)$ ) verwendet, die von den in einer Trainingsphase (1) verwendeten Trainingsdaten (x) abhängen.  
Durch Berücksichtigung der "Verlässlichkeit" jedes Schätzers (Si), die sich beispielsweise aus der Varianz ( $\text{Var}(Si(xg))$ ) jedes Schätzers (Si) ergibt, werden die Schätzungsergebnisse des Gesamtschätzers (GS) erheblich verbessert.



DE 195 26 954 C 1

## Beschreibung

Bei künstlichen neuronalen Netzen und anderen Schätzern, die auf statistischen Verfahren basieren, ist es oftmals sehr nützlich, wenn ein komplexes Problem einer Abbildung von einer Kombination von Eingangssignalen  $x \in \mathbb{R}^D$  auf eine Kombination von Ausgangssignalen  $y \in \mathbb{R}^K$  in eine Mehrzahl kleinerer Abbildungen aufgeteilt wird.

Dabei ist es bekannt, die Aufteilung eines komplexen rechnergestützten statistischen Schätzers, also z. B. eines künstlichen neuronalen Netzes, in eine Mehrzahl einfacherer rechnergestützte statistische Schätzer durchzuführen und die Ausgangssignale der Mehrzahl einfacherer rechnergestützter statistischer Schätzer zu gewichten und die gewichteten Ausgangssignale in einem Gesamtschätzer zu einem Ergebnissignal akkumulieren (R. Jacobs et al. Adaptive Mixtures of Local Experts, Neural Computation, Vol. 3, Massachusetts Institute of Technology, 1991, S. 79–87).

Dabei ist es bekannt, daß die Gewichtsfunktionen mit dem Training der einzelnen Schätzer bestimmt werden und somit an die Trainingsdaten angepaßt werden. Dabei wird das bekannte Expectation-Maximization Verfahren, EM-Verfahren genannt, zur Bildung der Gewichtsfunktionen verwendet, wobei die gesamte Mehrzahl der einfacheren Schätzer alle mit denselben Trainingsdaten trainiert werden.

Um die Ergebnisse bei der Akkumulierung zu verbessern, ist es bekannt, die Ausgangssignale aller Module, also aller einzelnen Schätzer zu mitteln, und somit den Gesamtschätzer zu verbessern (M. Perrone, Improving Regression Estimates: Averaging Methods for Variance Reduction with Extensions to General Convex Measure Optimization, PhD thesis, Brown University, USA, S. 10–21, 1993).

Die Tatsache, daß sich der Begriff eines statistischen Schätzers nicht nur auf künstliche neuronale Netze beschränkt, sondern alle rechnergestützten Schätzer umfaßt, die auf statistischen Verfahren basieren, ist jedem Fachmann bekannt und sofort ersichtlich.

Eine Übersicht über rechnergestützte Schätzer, die auf statistischen Verfahren basieren, ist in (J. Hartung et al., Statistik, Oldenbourg Verlag, München, 9. Auflage, ISBN 3-486-22055-1, S. 123–142, 1993) gegeben.

Das erfundungsgemäße Verfahren ist für alle rechnergestützten statistischen Schätzer einsetzbar. Weitere statistische Schätzer, die in (J. Hartung et al., Statistik, Oldenbourg Verlag, München, 9. Auflage, ISBN 3-486-22055-1, S. 123–142, 1993) nicht angegeben sind, aber für die das erfundungsgemäße Verfahren selbstverständlich auch einsetzbar ist, sind jedem Fachmann bekannt.

Es sind unterschiedliche Möglichkeiten zur Berechnung der Varianz eines einzelnen Moduls, also eines einzelnen Schätzers, bekannt (B. Efron et al., An Introduction to the Bootstrap, Chapman & Hall Verlag, New York, USA, ISBN 0-412-04231-2, S. 39–45, 1993; und R. Tibshirani, A Comparison of Some Error Estimates for Neural Network Models, Department of Preventive Medicine and Biostatistics, University of Toronto, 1994).

Aus (B. Efron et al., An Introduction to the Bootstrap, Chapman & Hall Verlag, New York, USA, ISBN 0-412-04231-2, S. 124–140, 1993) sind die Grundlagen über auftretende Verzerrungen der Schätzer, dem sogenannten Bias, und deren Schätzung zur deren Kompensation, bekannt.

Das bekannte Verfahren der Gewichtung der rechnergestützten statistischen "kleineren" Schätzern birgt vor allem den Nachteil einer erheblichen Ungenauigkeit in der Akkumulierung in dem Gesamtschätzer zu einem Ergebnissignal, dem Schätzungsresultat, in sich.

Somit liegt der Erfolg das Problem zugrunde, ein Verfahren anzugeben, das die Ungenauigkeit eines Gesamtschätzers, die durch Gewichtsfunktionen verursacht wird, die Ausgangssignale von "kleineren" rechnergestützten statistischen Schätzern gewichten und als Eingangssignale des Gesamtschätzers zur Verfügung stellen, verringert.

Das Problem wird durch das Verfahren gemäß Patentanspruch 1 gelöst.

Bei dem Verfahren werden Gewichtsfunktionen verwendet, die abhängig sind von der Verlässlichkeit der einzelnen Schätzer, die abhängig sind von den Trainingsdaten, mit denen die einzelnen "kleineren" rechnergestützten statistischen Schätzer in einer Trainingsphase trainiert wurden. Dadurch wird die Genauigkeit des Gesamtschätzers erheblich erhöht.

Dies wird vor allem dadurch erreicht, daß durch die Abhängigkeit der Gewichtsfunktionen von den verwendeten Trainingsdaten die Verlässlichkeit jedes einzelnen "kleineren" rechnergestützten statistischen Schätzers abhängig von den Trainingsdaten, mit denen jeder einzelne "kleinere" rechnergestützte statistische Schätzer berücksichtigt wird bei der Akkumulation der Ausgangssignale der "kleineren" rechnergestützten statistischen Schätzer in einer Generalisierungsphase in dem Gesamtschätzer zu dem Ergebnissignal für die angelegten Eingangssignale der "kleineren" rechnergestützten statistischen Schätzer.

Weiterbildungen der Erfahrung ergeben sich aus den abhängigen Ansprüchen.

Im folgenden wird die Erfahrung anhand von lediglich ein bevorzugtes Ausführungsbeispiel darstellenden Zeichnungen näher erläutert.

Es zeigen

Fig. 1 eine Skizze, in der schematisch die Anordnung dargestellt ist, in der Ausgangssignale vieler "kleinerer" statistischer Schätzer gewichtet werden, abhängig von den Trainingsdaten, mit denen die Schätzer trainiert wurden, und in einem Gesamtschätzer zu einem Ergebnissignal akkumuliert werden;

Fig. 2 eine Skizze, in der schematisch die Anordnung dargestellt ist, mit der Ausgangssignalen vieler "kleinerer" statistischer Schätzer gewichtet werden abhängig von den Trainingsdaten, mit denen die Schätzer trainiert wurden, wobei unterschiedliche Schätzer mit unterschiedlichen Trainingsdaten trainiert wurden, und in einem Gesamtschätzer zu einem Ergebnissignal akkumuliert werden;

Fig. 3 ein Ablaufdiagramm, das den Ablauf des erfundungsgemäßen Verfahrens darstellt;

Fig. 4 eine Skizze, in der Möglichkeiten dargestellt sind, wie man die Gewichtsfunktionen bildet.

Anhand der Fig. 1 bis 4 wird das erfundungsgemäße Verfahren weiter erläutert.

In Fig. 1 ist eine Anordnung dargestellt mit "kleineren" rechnergestützten statistischen Schätzern  $S_i$ , d. h. rechnergestützten Schätzern, die auf statistische Verfahren beruhen. Hierbei ist ein Index  $i$ , der jeden "kleineren" rechnergestützten statistischen Schätzern  $S_i$  eindeutig identifiziert, eine natürliche Zahl in dem Intervall  $[1 \dots n]$ .

Es ist also eine Anordnung mit einer beliebigen Anzahl von  $n$  "kleineren" rechnergestützten statistischen Schätzern  $S_i$  für eine Anordnung, für das das erfundungsgemäße Verfahren angewendet werden kann, vorgesehen.

Unter rechnergestützten statistischen Schätzern  $S_i$  sind in diesem Zusammenhang z. B. künstliche neuronale Netze, oder andere Schätzer, die auf statistischen Verfahren basieren, die z. B. aus (J. Hartung et al., Statistik, Oldenbourg Verlag, München, 9. Auflage, ISBN 3-486-22055-1, S. 123 – 142, 1993) bekannt sind, zu verstehen.

Im folgenden wird jede Parameteranpassung eines statistischen Schätzers  $S_i$  als Training des Schätzers  $S_i$  bezeichnet.

Die rechnergestützten statistischen Schätzer  $S_i$  sind Schätzer, die mit Hilfe einer Computereinheit CE realisiert werden. Aus diesem Grund sind die statistischen Schätzer  $S_i$  rechnergestützt. Die Tatsache, daß z. B. neuronale Netze und auch alle anderen Arten statistischer Schätzer nichts anderes darstellen als eine eigene Art von Datenverarbeitungsanlage, also eine eigene Art von Computereinheit, ist jedem Fachmann völlig geläufig.

In einer Trainingsphase 1 werden an die einzelnen "kleineren" rechnergestützten statistischen Schätzer  $S_i$  Trainingsdaten  $x$  angelegt, mit denen die Schätzer trainiert werden (vgl. Fig. 3).

Algorithmen zum Trainieren statistischer Schätzer sind jedem Fachmann hinlänglich bekannt, und können ohne Einschränkungen zum Trainieren der "kleineren" rechnergestützten statistischen Schätzer  $S_i$  eingesetzt werden. Eine Übersicht über eine Auswahl von Trainingsalgorithmen ist in (J. Hertz et al., Introduction to the Theory of Neural Computation, Addison-Wesley Publishing Company, ISBN 0-201-51560-1, S. 89 – 156, 1991) beschrieben.

Es ist ebenso vorgesehen, daß nicht alle "kleineren" rechnergestützten statistischen Schätzer  $S_i$  mit denselben Trainingsdaten  $x$  trainiert werden, sondern sie können auch, wie in Fig. 2 dargestellt, jeweils nur mit einem Teil  $tx_j$  der Trainingsdaten  $x$  trainiert werden. Hierbei ist ein Index  $j$ , der jeden Teil  $tx_j$  der Trainingsdaten  $x$  eindeutig identifiziert, eine beliebige natürliche Zahl im Intervall  $[1 \dots m]$ .

Die Gesamtheit der mehreren "kleineren" rechnergestützten statistischen Schätzer  $S_i$  wird im folgenden als eine Vielzahl VZ von Schätzern  $S_i$  bezeichnet.

Hierbei bezeichnen die Teile  $Tx_j$  der Trainingsdaten  $x$  nicht nur einzelne Muster, die zum Training der Vielzahl VZ der Schätzer  $S_i$  verwendet werden, sondern sie können ebenso einzelne Teile eines oder mehrerer zum Training verwendeter Muster umfassen.

Die Teile  $tx_j$  der Trainingsdaten  $x$  können sich in den Trainingsdaten  $x$  überschneiden, können aber auch disjunkt sein.

Es ist auch vorgesehen, daß jeweils ein Teil  $tx_j$  der Trainingsdaten  $x$  für genau einen "kleineren" rechnergestützten statistischen Schätzer  $S_i$  verwendet wird.

Es ist aber ebenso vorgesehen, daß in einer weiteren Ausführungsform ein Teil  $tx_j$  der Trainingsdaten  $x$  für mehrere "kleinere" rechnergestützte statistische Schätzer  $S_i$  verwendet werden oder mehrere Teile  $tx_j$  der Trainingsdaten  $x$  für einen "kleineren" rechnergestützten statistischen Schätzer  $S_i$  verwendet werden.

Durch Gewichtsfunktionen  $hi(xg)$  werden Ausgangssignale ASi der Vielzahl VZ von Schätzern  $S_i$  in einer Generalisierungsphase GP 3 gewichtet, wenn zu erkennende bzw. zu klassifizierende oder zu schätzende Daten ED, an die Vielzahl VZ von Schätzern  $S_i$  gelegt werden (vgl. Fig. 3).

Hierbei bezeichnet ein Datenpunkt  $xg$  ein Datum, das in der Generalisierungsphase GP an die Vielzahl VZ von Schätzern  $S_i$  angelegt wird.

Das Verfahren ist auch ohne Einschränkungen anwendbar auf an die Vielzahl VZ der Schätzer  $S_i$  angelegten Daten ED, die durch eine Regression durch die Vielzahl VZ der Schätzer  $S_i$  und den Gesamtschätzer GS an die Daten ED approximiert werden sollen.

Die Generalisierungsphase GP ist eine der Trainingsphase 1 zeitlich folgenden Phase, in der die Daten ED an die Vielzahl VZ der einzelnen Schätzer  $S_i$  angelegt werden. Hierbei müssen die Gewichtsfunktionen  $hi(xg)$  schon gebildet worden sein. Die Bildung der Gewichtsfunktionen  $hi(xg)$  wird in unterschiedlichen Ausführungsformen im weiteren beschrieben.

Dies bedeutet, daß gewichtete Ausgangssignale GASi der Vielzahl VZ von Schätzern  $S_i$  sich ergeben aus

$$GASi = hi(xg) \cdot ASi$$

Hierbei bezeichnet der Index  $i$  wiederum jeweils das Ausgangssignale ASi, die Gewichtsfunktionen  $hi(xg)$  und die gewichteten Ausgangssignale GASi eindeutig, die jeweils alle dem i-ten Schätzer  $S_i$  zugeordnet werden.

Die gewichteten Ausgangssignale GASi werden in einem Gesamtschätzer GS akkumuliert und bilden ein Ergebnissignal ES(xg). Das Ergebnissignal ES(xg) kann auch ein Vektor beliebiger vorgegebener Dimension sein.

Die Gewichtsfunktionen  $hi(xg)$  werden in der Trainingsphase 1 gebildet, abhängig von den Trainingsdaten  $x$  oder den Teilen  $tx_j$  der Trainingsdaten  $x$ , wie im weiteren beschrieben wird.

Im weiteren sind einige Varianten A beschrieben zur Bildung 2 der Gewichtsfunktionen  $hi(xg)$  (vgl. Fig. 3 und 4).

Die Gewichtsfunktionen  $hi(xg)$  werden abhängig von den Trainingsdaten  $x$  oder Teilen  $tx_j$  der Trainingsdaten  $x$  gebildet (vgl. Fig. 3).

Eine vorteilhafte Bildung der Gewichtsfunktionen  $hi(xg)$  für jeden Schätzer  $S_i$  geschieht durch Berücksichtigung der Varianz  $Var(Si(xg))$  B des jeweiligen Schätzers  $S_i$  für ein in der Generalisierungsphase GP an die Vielzahl VZ der Schätzer  $S_i$  angelegtes Datum  $xg$ .

Wie im vorigen beschrieben wurde, sind verschiedene Möglichkeiten zur Berechnung der Varianz  $\text{Var}(\hat{S}_i(xg))$  eines Schätzers  $\hat{S}_i$  bekannt (B. Efron et al., An Introduction to the Bootstrap, Chapman & Hall Verlag, New York, USA, ISBN 0-412-04231-2, S. 39–45, 1993; und R. Tibshirani, A Comparison of Some Error Estimates for Neural Network Models, Department of Preventive Medicine and Biostatistics, University of Toronto, 1994).

- 5 Die Varianz  $\text{Var}(\hat{S}_i(xg))$  eines Schätzers  $\hat{S}_i$  wird für die jeweilige Gewichtsfunktion  $h_i(xg)$  berücksichtigt, indem die Gewichtsfunktion  $h_i(xg)$  gebildet wird aus dem Kehrwert der Varianz  $\text{Var}(\hat{S}_i(xg))$  eines Schätzers  $\hat{S}_i$ , also ergibt sich jede Gewichtsfunktion  $h_i(xg)$  aus:

$$10 \quad h_i(xg) = \frac{1}{\text{Var}(\hat{S}_i(xg))} \quad (\text{vgl. Figur 4}).$$

- 15 Diese Art B der Bildung der Gewichtsfunktionen  $h_i(xg)$  eignet sich besonders für den Fall, daß die einzelnen Schätzer  $\hat{S}_i$  der Vielzahl VZ von Schätzern  $\hat{S}_i$  in der Trainingsphase 1 mit unterschiedlichen Teilen  $t_{xj}$  der Trainingsdaten  $x$  trainiert werden, wobei die unterschiedlichen Teile  $t_{xj}$  der Trainingsdaten  $x$  aber ein und dasselbe Eingangs-Ausgangs-Verhalten einer Abbildung beschreiben.

- 20 Anschaulich bedeutet diese Art der Bildung der Gewichtsfunktionen  $h_i(xg)$ , daß die Ausgangssignale  $A\hat{S}_i$  eines unsicheren Schätzers, also eines Schätzers  $\hat{S}_i$  der Vielzahl VZ von Schätzern  $\hat{S}_i$ , der an dem jeweiligen Datenpunkt  $xg$  eine große Varianz  $\text{Var}(\hat{S}_i(xg))$  aufweist, geringer gewichtet werden, als die Ausgangssignale  $A\hat{S}_i$  eines sichereren Schätzers  $\hat{S}_i$ , also eines Schätzers  $\hat{S}_i$ , der an dem jeweiligen Datenpunkt  $xg$  eine geringere Varianz  $\text{Var}(\hat{S}_i(xg))$  aufweist.

- 25 Die Tatsache, daß die im vorigen beschriebene Art der Bildung der Gewichtsfunktionen  $h_i(xg)$  für den Fall des Trainings mit unterschiedlichen Teilen  $t_{xj}$  der Trainingsdaten  $x$  für jeden Schätzer  $\hat{S}_i$  der Vielzahl VZ von Schätzern  $\hat{S}_i$  vorteilhaft einsetzbar ist, bedeutet aber in keinem Fall, daß diese Art der Bildung der Gewichtsfunktionen  $h_i(xg)$  nicht auch für alle anderen Fälle, also z. B. für das Training der Vielzahl VZ von Schätzern  $\hat{S}_i$  mit den gleichen Trainingsdaten  $x$  einsetzbar ist.

- 30 Eine weitere Möglichkeit, die Gewichtsfunktionen  $h_i(xg)$  zu bilden, liegt darin, die Wahrscheinlichkeit  $P(i|xg)$ , also die Wahrscheinlichkeit, daß der Datenpunkt  $xg$  zu einem bestimmten Schätzer  $\hat{S}_i$  gehört, in den Gewichtsfunktionen  $h_i(xg)$  zu berücksichtigen D (vgl. Fig. 4).

Also ergeben sich die Gewichtsfunktionen  $h_i(xg)$  aus D:

$$35 \quad h_i(xg) = P(i|xg) = \frac{P(i, xg)}{P(xg)} \propto P(xg|i) \cdot P(i).$$

Die Wahrscheinlichkeit  $P(i|xg)$  kann auf verschiedenen Wegen geschätzt werden, die jedem Fachmann bekannt sind.

- 40 Eine vorteilhafte Möglichkeit zur Schätzung der Wahrscheinlichkeit  $P(xg|i)$  besteht in der Verwendung eines Gauß-Modells:

$$45 \quad P(xg|i) = \sum_{l=1}^o P^{il} D(xg; \text{Par}^{il}).$$

Hierbei bezeichnet der Term  $D(xg; \text{Par}^{il})$  eine mehrdimensionale lokale DichteVerteilung einem Parametersatz  $\text{Par}^{il}$ .

- 50 Jeder Schätzer  $\hat{S}_i$  weist o unterschiedliche lokale DichteVerteilungen  $D(xg; \text{Par}^{il})$  auf, die jeweils durch einen Index l eindeutig identifiziert werden, die in

Ein weiterer Parametersatz  $P^{il}$  beschreibt die "Stärke" der einzelnen lokalen DichteVerteilung  $D(xg; \text{Par}^{il})$  an dem Datenpunkt  $xg$  für den Schätzer  $\hat{S}_i$ .

- 55 In einem speziellen Fall kann der Term auch mit einer mehrdimensionalen Gauß-NormalVerteilung realisiert werden. In diesem Fall ergibt sich für:

$$D(xg; \text{Par}^{il}) = G(xg; c^{il}; \Sigma^{il}).$$

- 60 Hierbei bezeichnet  $G(xg; c^{il}; \Sigma^{il})$  eine Gaußsche NormalVerteilung mit einem Zentrumsvektor  $c^{il}$  und einer Kovarianzmatrix  $\Sigma^{il}$  an dem Datenpunkt  $xg$ .

Der weitere Parametersatz  $P^{il}$  sowie der Parametersatz  $\text{Par}^{il}$  können beispielsweise durch den bekannten EM-Algorithmus bestimmt werden (A. Dempster et al., Maximum Likelihood from incomplete data via the EM algorithm, Journal of the Royal Statistical Society B, 1989).

- 65 Weitere Möglichkeiten zur Bestimmung des weiteren Parametersatzes  $P^{il}$  sowie des Parametersatzes  $\text{Par}^{il}$  sind jedem Fachmann geläufig und können ohne Einschränkungen eingesetzt werden.

Diese im vorigen beschriebene Art der Bildung der Gewichtsfunktionen  $h_i(xg)$  ist besonders geeignet, jedoch keineswegs beschränkt auf den Fall, daß die einzelnen Schätzer  $\hat{S}_i$  der Vielzahl VZ von Schätzern  $\hat{S}_i$  in der Trainingsphase 1 mit unterschiedlichen Teilen  $t_{xj}$  der Trainingsdaten  $x$  trainiert wurden, wobei die unterschiedlich-

chen Teile  $t_{ij}$  der Trainingsdaten  $x$  unterschiedliches Eingangs-Ausgangs-Verhalten eines Schätzers  $S_i$ . Das Ergebnissignal  $ES(xg)$  ergibt sich in den im vorigen beschriebenen Fällen, für jeden Fachmann bekannt, in der folgenden Weise:

$$ES(xg) = \frac{1}{\sum_{i=1}^n h_i(xg)} \cdot \sum_{i=1}^n h_i(xg) \cdot S_i(xg)$$

Für die bisher beschriebenen Arten, die Gewichtsfunktionen  $h_i(xg)$  zu bilden, wird angenommen, daß die Schätzer  $S_i$  voneinander unabhängig sind.

Für den Fall, daß die einzelnen Schätzer  $S_i$  der Vielzahl VZ von Schätzern  $S_i$  nicht voneinander unabhängig sind, ist eine weitere Möglichkeit geeignet, das Ergebnissignal  $ES(xg)$  zu bilden C (vgl. Fig. 4).

Dies schließt die Anwendung der im folgenden beschriebenen Vorgehensweise zur Bestimmung des Ergebnissignals  $ES(xg)$  auf unabhängige Schätzer  $S_i$  keineswegs aus.

Zur anschaulicherer Erläuterung dieser Verfahrensvariante werden einige Annahmen getroffen, die allerdings nicht die Allgemeingültigkeit der im folgenden beschriebenen Variante einschränken, wie ebenso im folgenden dargestellt beschrieben wird.

Es werden zur Veranschaulichung zwei Schätzer  $S_1$  und  $S_2$  angenommen, deren Abhängigkeit durch folgende Beziehung gegeben ist:

$$P(f_1(xg), f_2(xg)) \propto \exp\left(-\frac{1}{2 \cdot \text{Var}_C} (f_1(xg) - f_2(xg))^2\right).$$

Im folgenden bezeichnen Funktionen  $S_i(xg)$  die Abbildungen der Schätzer  $S_i$  an dem Datenpunkt  $xg$ , also die "Vorhersagen" der Schätzer  $S_i$ , die gebildet werden aus der eigentlichen Abbildung  $f_i(xg)$  und einem unabhängigen Gaußschen Rauschen  $\epsilon_i$ , dem die jeweiligen Abbildungen  $f_i(xg)$  unterliegen, mit der Varianz  $\text{Var}(S_i(xg))$  eines einzelnen Schätzers  $S_i$  der Vielzahl VZ von Schätzern  $S_i$ :

$$S_i(xg) = f_i(xg) + \epsilon_i$$

Die Funktionen  $f_1(x)$  und  $f_2(x)$  werden geschätzt, also angenähert durch:

$$f_1(xg) = \frac{1}{K(xg)} ((\text{Var}(S_2(xg)) + \text{Var}_C) \cdot S_1(xg) + \text{Var}(S_1(xg)) \cdot S_2(xg))$$

$$f_2(xg) = \frac{1}{K(xg)} ((\text{Var}(S_2(xg)) \cdot S_1(xg) + (\text{Var}(S_1(xg)) + \text{Var}_C) \cdot S_2(xg))$$

$$K(xg) = \text{Var}(S_1(xg)) + \text{Var}(S_2(xg)) + \text{Var}_C$$

Eine Kopplungsvarianz  $\text{Var}_C$  gibt die Stärke der Kopplung zwischen den beiden Schätzern  $S_i$ , also die Abhängigkeiten der Schätzer  $S_1$  und  $S_2$  an.

Je größer die Kopplungsvarianz  $\text{Var}_C$  ist, desto weniger sind die Schätzer  $S_1$  und  $S_2$  "gekoppelt".

Geht also der Wert der Kopplungsvarianz  $\text{Var}_C \rightarrow \infty$ , so sind die beiden Schätzer  $S_1$  und  $S_2$  "entkoppelt", also voneinander unabhängig.

Wenn die Kopplungsvarianz  $\text{Var}_C$  den Wert  $\text{Var}_C = 0$  aufweist, so ist das Eingangs-Ausgangs-Verhalten der Schätzer  $S_1$  und  $S_2$  identisch.

Das Ergebnissignal  $ES(xg)$  ergibt sich dann in der folgenden Weise:

$$ES(xg) = P(1|xg) \cdot f_1(xg) + P(2|xg) \cdot f_2(xg)$$

Der an zwei Schätzern  $S_1$  und  $S_2$  veranschaulichte Fall der Abhängigkeit der Schätzer  $S_i$  voneinander wird verallgemeinert in dem folgenden Ausdruck:

$$f_i(xg) = \frac{1}{K(xg)} \left( \frac{S_i(xg)}{\text{Var}(S_i(xg))} + \sum_{k \neq i} \frac{S_k(xg)}{\text{Var}(S_k(xg)) + \text{Var}_C^{ik}} \right)$$

5

$$K(xg) = \frac{1}{\text{Var}(S_i(xg))} + \sum_{k \neq i} \frac{1}{\text{Var}(S_k(xg)) + \text{Var}_C^{ik}}$$

10

Für den allgemeinen Fall wird folgende Abhängigkeit der Schätzer  $S_i$  angenommen:

$$P(f_1(xg), f_2(xg), \dots, f_n(xg)) \propto \exp\left(-\frac{1}{2} \sum_{i=1}^n \sum_{k=1}^n \frac{1}{\text{Var}_C^{ik}} \cdot (f_i(xg) - f_k(xg))^2\right)$$

Für den Fall, daß nur ein Trainingsdatensatz für die Vielzahl VZ von Schätzern  $S_i$  vorliegt, also die Trainingsdaten  $x$ , kann die Gesamtschätzung noch weiter verbessert werden, indem die Gewichtsfunktionen  $h_i(xg)$  auf folgende Art E gebildet werden (vgl. Fig. 4):

$$h_i(xg) = \frac{1}{\text{Var}(S_i(xg)) + \text{Res}(S_i(xg))}$$

25

Hierbei ist mit  $\text{Res}(S_i(xg))$  ein durch statistisches Rauschen verursachter Restfehler bezeichnet.

Der Restfehler  $\text{Res}(S_i(xg))$  kann geschätzt werden, indem ein Trainingsfehler in der Trainingsphase 1, der durch statistisches Rauschen verursacht wird, beispielsweise mit einem neuronalen Netz approximiert wird.

In den bisherigen Fällen wurde angenommen, daß die einzelnen Schätzer  $S_i$  untereinander nicht korreliert sind.

Für den Fall, daß die einzelnen Schätzer  $S_i$  in der Trainingsphase 1 nicht mit völlig unterschiedlichen Trainingsdaten  $x$  trainiert wurden, d. h., daß sich die Teile  $tx_j$  der Trainingsdaten  $x$ , mit denen die einzelnen Schätzer  $S_i$  der Vielzahl VZ von Schätzern  $S_i$  zum mindesten zum Teil überschneiden, können die einzelnen Schätzer  $S_i$  der Vielzahl VZ von Schätzern  $S_i$  korreliert sein.

Eine Kovarianzmatrix  $\Omega(xg)$  der Dimension  $n \times n$  beschreibt existierende Kovarianzen zwischen den einzelnen Schätzern  $S_i$  der Vielzahl VZ von Schätzern  $S_i$  für den Datenpunkt  $xg$ .

Ein  $n$ -dimensionaler Vektor aller Gewichtsfunktionen  $h_i(xg)$  ergibt sich in diesem Fall aus:

$$40 \quad h(xg) = \Omega^{-1}(xg) \cdot u$$

Hierbei bezeichnet ein Vektor  $u$  einen  $n$ -dimensionalen Einheitsvektor.

Für den Fall, daß die einzelnen Schätzer  $S_i$  der Vielzahl VZ von Schätzern  $S_i$  Verzerrungen, dem sogenannten Bias, unterliegen, ergibt sich der  $n$ -dimensionale Vektor aller Gewichtsfunktionen  $h_i(xg)$  aus:

$$45 \quad h(xg) = (\Omega(xg) + B(xg))^{-1} \cdot u$$

Hierbei bezeichnet eine Bias-Matrix  $B(xg)$  der Dimension  $n \times n$  eine Matrix, in der Verzerrungen  $\text{bias}_i(xg)$  der einzelnen Schätzer  $S_i$  der Vielzahl VZ von Schätzern  $S_i$  eingetragen sind.

Nicht-Diagonalelemente  $B_{st}(x)$  der Bias-Matrix  $B(x)$  ergeben sich aus:

$$B_{st}(x) = \text{bias}_s(x) \cdot \text{bias}_t(x)$$

mit  $s \neq t$

55

#### Patentansprüche

1. Verfahren zur rechnergestützten Kombination einer Vielzahl (VZ) von Schätzern ( $S_i; i = 1 \dots n$ ), die auf statistischen Verfahren beruhen, zu einem Ergebnissignal ( $ES(xg)$ ) in einem Gesamtschätzer (GS),

60     – bei dem in einer Trainingsphase (1) die Schätzer ( $S_i$ ) der Vielzahl (VZ) von Schätzern ( $S_i$ ) mit Trainingsdaten ( $x$ ) oder Teilen ( $tx_j; j = 1 \dots m$ ) trainiert werden,

– bei dem für jeden Schätzer ( $S_i$ ) eine Gewichtsfunktion ( $h_i(xg); i = 1 \dots n$ ) gebildet wird, die das Gewicht festlegt, mit dem jeder Schätzer ( $S_i$ ) jeweils bei der Kombination in dem Gesamtschätzer (GS) zu dem Ergebnissignal ( $ES(xg)$ ) berücksichtigt wird, und

65     – bei dem jede Gewichtsfunktion ( $h_i(xg)$ ) von den Trainingsdaten ( $x$ ) oder den Teilen ( $tx_j$ ) der Trainingsdaten ( $x$ ) der einzelnen Schätzer ( $S_i$ ) der Vielzahl (VZ) von Schätzern ( $S_i$ ) abhängt.

2. Verfahren nach Anspruch 1, bei dem die Gewichtsfunktion ( $h_i(xg)$ ) verwendet wird, die sich ergibt aus

$$h_i(xg) = \frac{1}{\text{var}(S_i(xg))}$$

wobei der Term  $\text{Var}(S_i(xg))$  die Varianz des Schätzers ( $S_i$ ) für einen Datenpunkt ( $xg$ ) bezeichnet.

3. Verfahren nach Anspruch 1, bei dem die Gewichtsfunktion ( $h_i(xg)$ ) verwendet wird, die sich ergibt aus 5

$h_i(xg) = P(i|xg)$ , wobei der Term  $P(i|xg)$  die Wahrscheinlichkeit angibt, daß ein Datenpunkt ( $xg$ ) einem Schätzer ( $S_i$ ) zugeordnet wird.

4. Verfahren nach Anspruch 1, bei dem bei voneinander abhängigen Schätzern ( $S_i$ ) der Vielzahl (VZ) von 10

$$f_i(xg) = \frac{1}{K(xg)} \left( \frac{S_i(xg)}{\text{Var}(S_i(xg))} + \sum_{k \neq i} \frac{S_k(xg)}{\text{Var}(S_k(xg)) + \text{Var}_c^{ik}} \right)$$
15

$$K(xg) = \frac{1}{\text{Var}(S_i(xg))} + \sum_{k \neq i} \frac{1}{\text{Var}(S_k(xg)) + \text{Var}_c^{ik}}$$
20

5. Verfahren nach Anspruch 1, bei dem die Gewichtsfunktion ( $h_i(xg)$ ) verwendet wird, die sich ergibt aus 25

$$h_i(xg) = \frac{1}{\text{var}(S_i(xg)) + \text{Res}(S_i(xg))}$$

wobei der Term  $\text{Var}(S_i(xg))$  die Varianz des Schätzers ( $S_i$ ) für den Datenpunkt ( $xg$ ) bezeichnet und wobei 30

der Term  $\text{Res}(S_i(xg))$  einen durch statistisches Rauschen des Schätzers ( $S_i$ ) verursachten Restfehler des Schätzers ( $S_i$ ) bezeichnet.

Hierzu 4 Seite(n) Zeichnungen

35

40

45

50

55

60

65

**- Leerseite -**

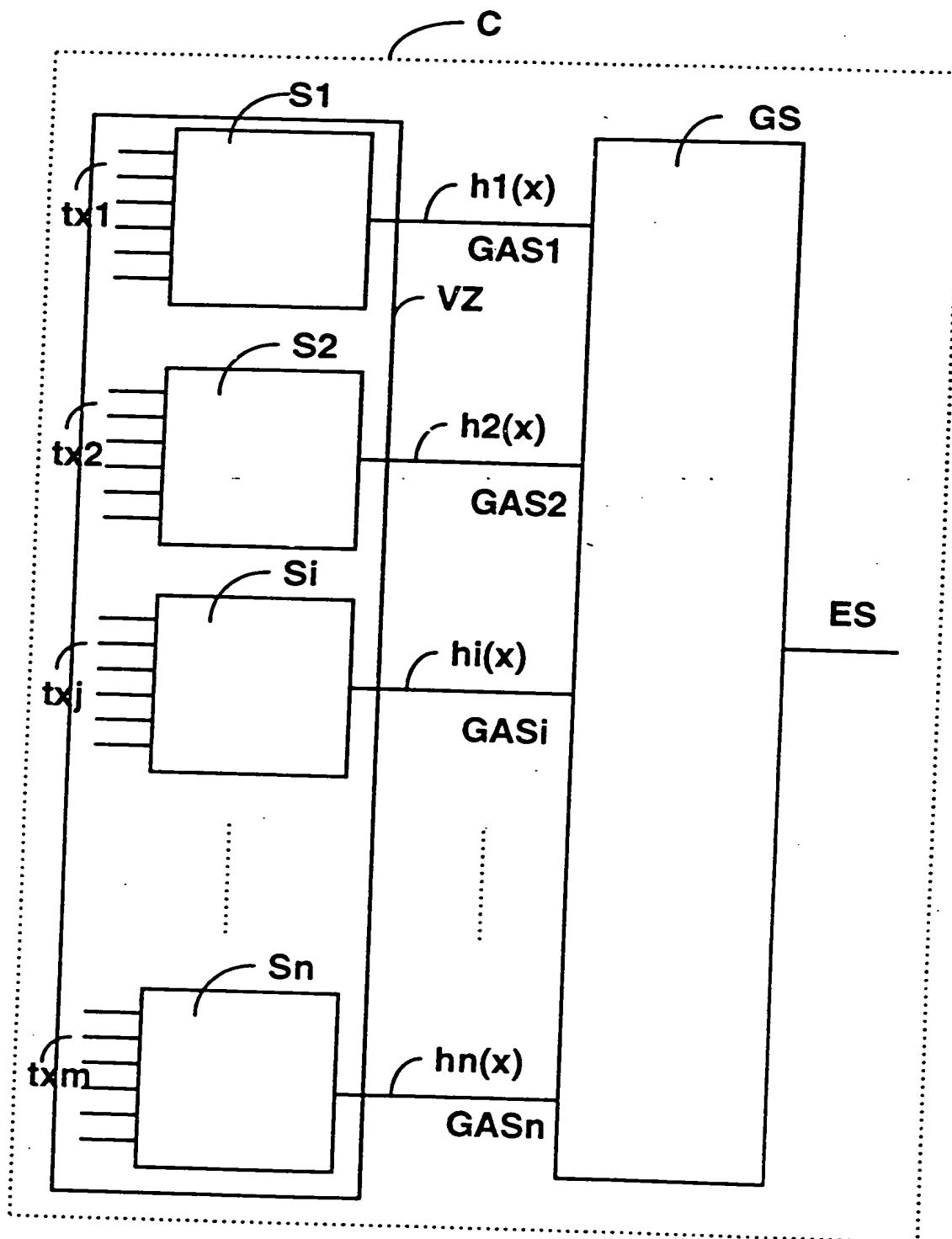
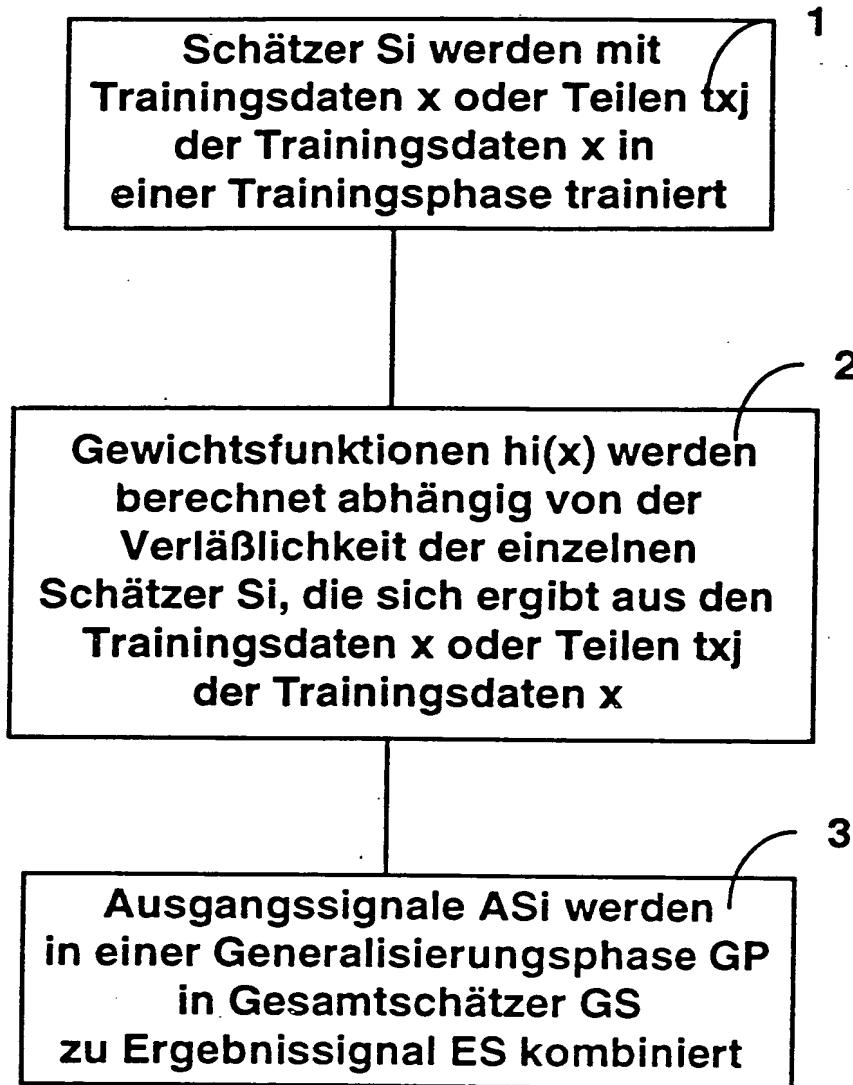


Fig. 2



**Fig. 3**

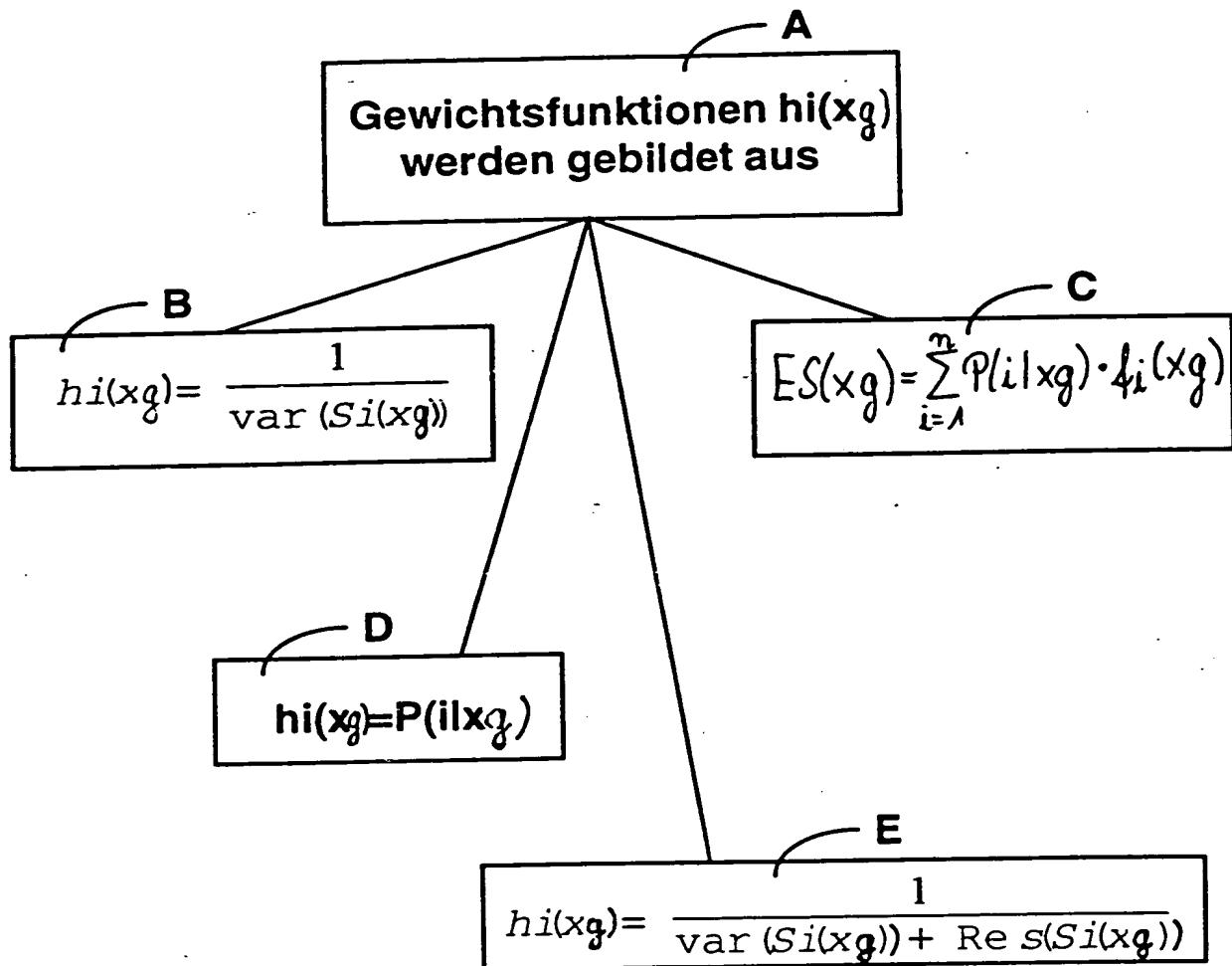


Fig. 4

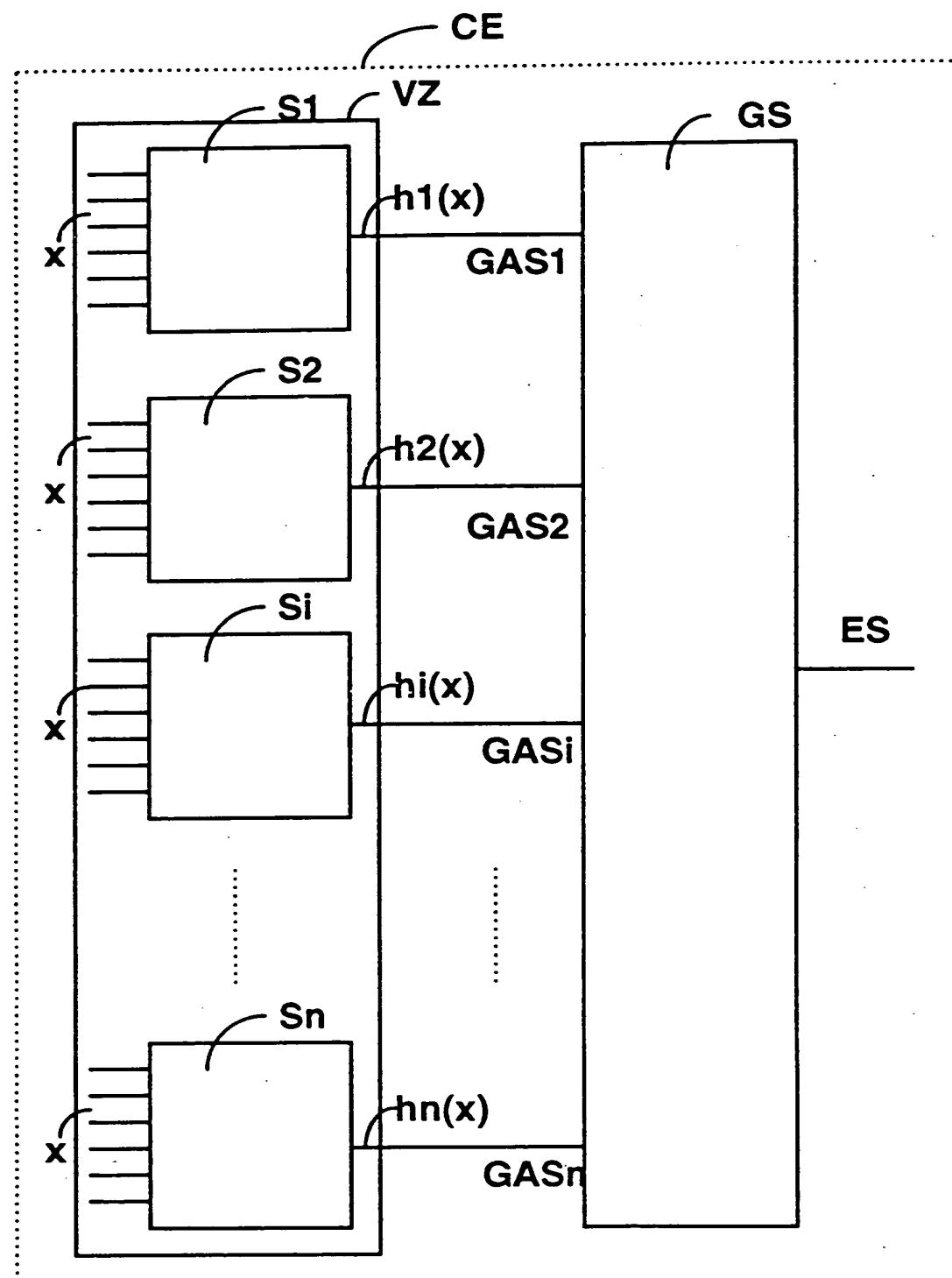


Fig. 1